

IN THE SPECIFICATION:

Please insert the following paragraphs beginning at page 4, line 2.

5 A permutation pattern is a pattern where the characters in the pattern can be in any order. For instance, in the input string, , a permutation pattern can be described as and the permutation pattern occurs at locations 1 and 7 in the input string. Permutation patterns have a variety of practical uses.

10 For example, genes that appear together consistently across genomes are believed to be functionally related: these genes in each other's neighborhood often code for proteins that interact with one another, suggesting a common functional association. However, the order of the genes in the chromosomes may not be the same. In other words, a group of genes appear in different permutations in the genomes. For example in plants, the majority of snoRNA genes are organized in polycistrons and transcribed as polycistronic precursor snoRNAs. Also, the olfactory receptor(OR)-gene superfamily is
15 the largest in the mammalian genome. Several of the human OR genes appear in clusters, with ten or more members located on almost all human chromosomes. Furthermore, some chromosomes contain more than one cluster, where a cluster has one or more permutation patterns.

20 As the available number of complete genome sequences of organisms grows, it becomes a fertile ground for investigation along the direction of detecting gene clusters by comparative analysis of the genomes. A gene *G* is compared with its orthologs in the different organism genomes. Even phylogenetically close species are not immune from gene shuffling, such as in *Haemophilus influenzae* and *Escherichia Coli*. Also, a multicistronic gene cluster sometimes results from horizontal transfer between
25 species and multiple genes in a bacterial operon fuse into a single gene encoding multi-domain protein in eukaryotic genomes.

If the functions of genes, say , are known, the function of its corresponding ortholog clusters may be predicted. Such positional correlation of genes as clusters and their corresponding orthologs have been used to predict functions of ABC

transporters and other membrane proteins.

The local alignment of nucleic or amino acid sequences, called the multiple sequence alignment problem, is based on similar subsequences; however the local alignment of genomes is based on detecting locally conserved gene clusters. A measure of gene similarity is used to identify the gene orthologs. For example, genes may be aligned with , and such an alignment is never detected in subsequence alignments.

Domains are portions of the coding gene (or the translated amino acid sequences) that correspond to a functional sub-unit of the protein. Often, these are detectable by conserved nucleic acid sequences or amino acid sequences. The conservation helps in a relative easy detection by automatic motif discovery tools. However, the domains may appear in a different order in the distinct genes giving rise to distinct proteins. But, they are functionally related due to the common domains. Thus these represent functionally coupled genes such as forming operon structures for co-expression.

15

Please amend the paragraph beginning at page 4, line 7, as follows.

The present invention allows permutation patterns to be discovered. In this disclosure, the ~~abstract~~ problem of discovering permutation patterns is formed as a discovery problem called the pattern problem and techniques that automatically discover permutation patterns in, for instance, multiple input patterns are given. As there is generally not enough knowledge about forming an appropriate model to filter the meaningful from the apparently meaningless permutation patterns, a model-less approach is taken herein, which allows all permutation patterns that appear a number of times to be determined. Additionally, a notation is introduced for maximal permutation patterns that drastically reduces the number of valid cluster patterns, without any loss of information, making it easier to study the results from an application viewpoint.

20
25